

Visualized Preference Transition Network Based on Recency and Frequency

Farid Masruri

Department of Computer Science and Intelligent System
Osaka Prefecture University, Osaka, Japan
E-mail: farid@mis.cs.osakafu-u.ac.jp

Hiroshi Tsuji[†]

Department of Computer Science and Intelligent System
Osaka Prefecture University, Osaka, Japan
E-mail: tsuji@cs.osakafu-u.ac.jp

Ryosuke Saga

Department of Information and Computer Sciences
Kanagawa Institute of Technology, Atsugi, Japan
E-mail: saga@ic.kanagawa-it.ac.jp

Received, March 17, 2011; Revised, August 31, 2011; Accepted, September 1, 2011

Abstract. Given a directed graph, we can determine how the user's preference moves from one product item to another. In this graph called "preference transition network", each node represents the product item while its edge pointing to the other nodes represents the transition of user's preference. However, with the large number of items make the network become more complex, unclear and difficult to be interpreted. In order to address this problem, this paper proposes a visualization technique in preference transition analysis based on recency and frequency. By adapting these two elements, the semantic meaning of each item and its transition can be clearly identified by its different types of node size, color and edge style. The experiment in a sales data has shown the results of the proposed approach.

Keywords: Data Mining, Information Visualization, Preference Analysis, Graph Theory, RFM Analysis

1. INTRODUCTION

In recent years, data mining has been widely used to discover and extract some important knowledge from large amount of data. Numerous methods and applications in data mining have been developed to tackle the information overload problem. Information visualization is considered as one of data mining approach (Han, 2006). The basic idea of visual data exploration is to present the data in some visual form, for example in a graph, allowing the human to get insight into the data, draw conclusions, and directly interact with the data. Visual data mining techniques have proven to be of high value in exploratory data analysis and they also have a high potential for exploring large databases (Keim, 2002).

In graph visualization, several techniques include using Key-graph (Ohsawa *et al.*, 1998) and FACT-Graph (Saga *et al.*, 2008) had been proposed. While Key-graph

has focused on co-occurrence among topics, FACT-Graph has focused on recency as well as frequency for expressing trends (Han and Kamber, 2006). Visualizing techniques in FACT-Graph are utilizing Graphviz as drawing engine.

Preference transition analysis is one of data mining techniques, which analyze user's preferences on choosing products or items in transaction data (Hayashi *et al.*, 2009). Its transition means that user's taste is always changing time by time. By visualizing it in a graph network called a "preference transition network", where the products or items are represented by nodes and its transitions are represented by edges, the method could help analysts to understand relations for choosing product items.

However, sometimes there are too many nodes and arrows in the generated network and it makes the visualized result ineffective for the analysis. Previously, the

[†] : Corresponding Author

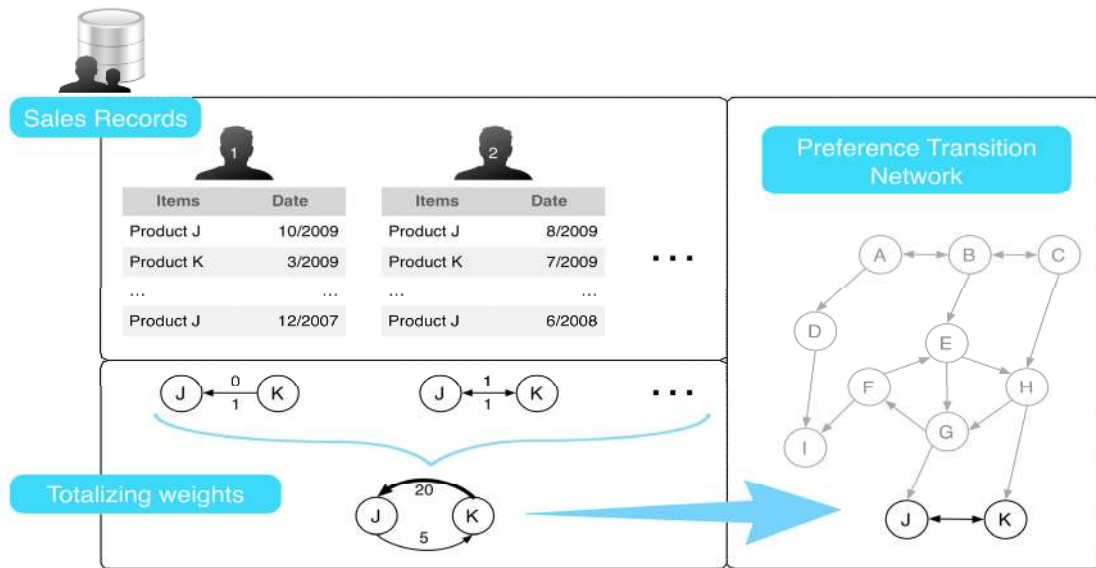


Figure 1. Concept of Preference Transition Network.

authors have proposed a simplification method on preference transition network in order to enhance and simplify the complex network (Hayashi *et al.*, 2010). This method worked based on directed acyclic graph theory by grouping cycles that appeared in graph so that the visualization became much clearer.

On the other hand, RFM (Recency/Frequency/Monetary) analysis is a marketing technique used to determine quantitatively which customers are the best ones by examining how recently a customer has purchased (*recency*), how often the customer purchase (*frequency*), and how much the customer spends (*monetary*). This paper proposes an approach to enhance the preference transition network, which is adapted from the concept of RFM analysis. Based on this scheme, the recency and frequency of all items are computed and afterwards the semantic meaning of each item and its transitions can be determined.

2. PREVIOUS WORKS and LIMITATIONS

2.1 Preference Transition Network

The basic concept of preference transition network is work based on graph theory. Suppose that there is a directed graph or digraph $G = (V, E)$ contains of:

- a set V , whose elements are called vertices or nodes ,
- a set E of ordered pairs of vertices called arcs directed edges,

then the customer movements or trends of customer on choosing product items can be visualized by using this

graph network. In this directed graph, each node V is representing product item, while its directed edges E representing the transition. Moreover, to construct such network, it is necessary to determine the direction of the edges by performing Mann-Whitney U test (Aczel *et al.*, 2005) of each transition or edge. Detailed scheme of preference network generation is shown in Figure 1.

The preference transition network had been used in several applications, such as a hotel recommender system (Saga *et al.*, 2008) and competitive relationships among product items analysis (Hayashi *et al.*, 2009). However, the previous method generates preference transition network in a very complex visualization. There are many nodes and edges, and it seems difficult to observe on which nodes the preference transitions concentrate and understand the aspect of the preference transitions among nodes. Hayashi *et al.* (2010) also had proposed a simplification method based on directed acyclic graph to tackle this problem, by grouping and eliminating equivalent nodes to make clearer visualization. However, it still suffers from two main problems: how to detect preference trend and how to clarify its transition. That is, the user's transitions on choosing product items were not taken into account and made it hard to be interpreted by the decision maker.

2.2 RFM Analysis

In business and marketing field, RFM analysis has widely used by direct marketers of all types for selecting which customers to target offers to. The basic principle underlying RFM analysis is that customers, who have purchased recently, have made more purchases and have

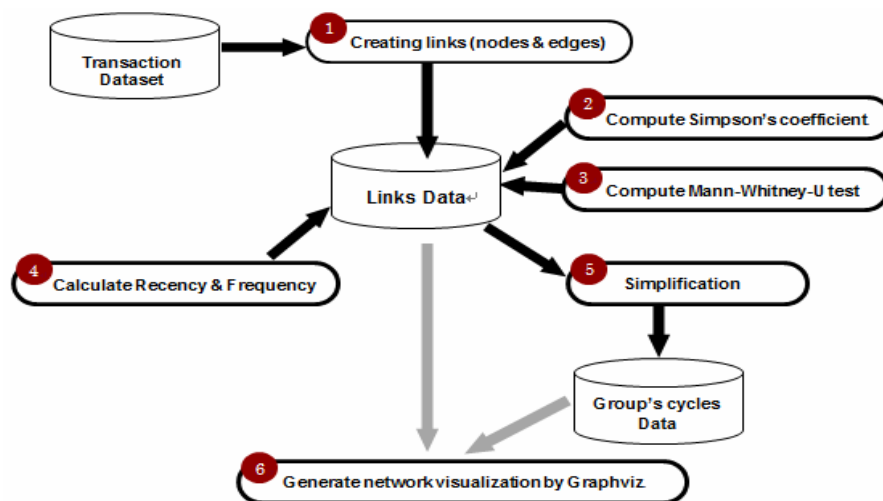


Figure 2. Process of Preference Transition Network Generation.

made larger purchases, are more likely to respond to the offering than other customers who have purchased less recently, less often and in smaller amounts (Miglautsch, 2002).

As an example shown in Table 1, customers are being segmented by a recency score based on date of most recent purchase or time interval since most recent purchase. This score is based on a simple ranking of recency values into a small number of categories. In a similar fashion, customers are then assigned a frequency ranking, with higher values representing a higher frequency of purchases. As in the example, which shows a three-category ranking scheme of recency and frequency, customers who purchases most often and most recent is classified as “Loyal Customers.”

Table 1. An example of Recency and Frequency Analysis.

	F: High	F: Intermediate	F: Low
R: Recently	Loyal Customers	Hopeful Customers	Probationary Customers
R: Intermediate	Frustrated Customers	Customers who have compelling reasons	
R: Past	Distrust Customers	Customers of other companies	

Furthermore, there are some several researches and applications which have adapted the concept of RFM, including visualization of newspaper’s trend (Terachi *et al.*, 2008) and customer segmentation based on RFM cluster (Aggelis, 2005). The first mentioned research used the idea of RFM analysis to propose a method of analyzing relationships between multiple words and analyzing keywords that appear infrequently by examining the transitions between keyword classes and changes in co-occurrence relationships. The last mentioned research has applied clustering algorithm to build a cus-

tom segmentation model for bank.

Whereas the RFM analysis is considered as simple yet powerful marketing tool and approach originally developed for database marketing, the RFM analysis also has several drawbacks as pointed by D’Auria (2009):

- RFM analyses say nothing about what is driving customers to make purchases.
- As RFM analysis is not a statistical modeling technique, it fails to predict the probability of a customer to respond to various marketing stimuli.
- RFM analysis cannot be used to predict changes in customer behavior over time, which is particularly relevant during periods of recession.

Hence, in this research, the simplicity and advantage of RFM analysis can be adapted to address the complexity and transition semantics in preference network’s visualization. By introducing RFM model into preference transition analysis, there is a research opportunity to improve its visualization. It is, therefore, this research aims to enhance the visualization of preference transition network by combining recency and frequency analysis and to give semantic meaning on its transitions.

3. PROPOSAL METHOD

3.1 Preprocessing Stage

First of all, in order to attain the objective, there are six preprocessing stages involved in generating a visualization of preference network, as shown in Figure 2. Previously, Hayashi *et al.* (2001) has proposed the step of preference transition network, covering from item pairs or links creation, Simpson’s coefficient computation and Mann-Whitney U-test. The proposal method still

covers and complements some basic steps with recency-frequency computation and its simplification method.

Additionally, there are two tables, named links table and cycle table, designed to maintain the computed item pairs in this preprocessing stage and another one to store group or cycle that occurs during the simplification step.

The initial stage is data extraction from transaction dataset to generate pairs of correlated items. In this step, the purchase history of each customer is being combined into a pair of items and every item pair is stored in links table. Then the second step involves co-occurrence measurement using Simpson’s coefficient. Suppose there is an item pair of X and Y, then its Simpson’s value is expressed as:

$$\text{Simpson}(X, Y) = \frac{\text{count}(X \cap Y)}{\min(\text{count}(X), \text{count}(Y))} \quad (1)$$

Until this step, preference transition network have been connected, but the direction of transition of every item pair is still unclear. Therefore, in order to identify this direction, the next step is applying Mann-Whitney U test on every item pair.

If the results of Mann-Whitney U test demonstrate that preference transition between two items is not distinct, that is to say, there is no notable difference in user’s preference in the item pair, and the edge between the two items is ignored. On the contrary, there is notable difference in user’s preference then adding direction to edge in order to show the flow of preference transition. Furthermore, the detailed method of recency and frequency is described in the next section.

3.2 RF-Table

As described in the previous section, to adapt the two elements of RFM analysis, that is recency and frequency, a so called RF-table has been built, as shown in Table 2. An RF-table defines the segmentation of an item and clarifies the semantic meaning of the item based on its RF score. Here, the recency score is determined from the purchase date of the item, while the frequency score is resolved from the number of transaction of the current item. In this way, there are 5×5 categories or partitions in RF-table containing seven segments.

Table 2. RF-Table for Item Segmentation.

		Frequency (the number of transactions)				
		5	4	3	2	1
Recency (item timestamp)	5	A		B		C
	4	Hot item		Appearing item		New item
	3	D		F		
	2					
	1	E		G		
Old item		Bad				

For example, if an item is recently released and has a large number of transactions then it is categorized as class A, which means that this particular item is popular. When an item has high frequency but it has very low recency score then it is categorized as class E, considered as an old item.

Note that the interval partition or discretization of recency and frequency category can be done in several ways, but basically there are general types of data partitioning: equal-width and equal-depth partitioning. In equal width or distance partitioning simply divides the range into N intervals of equal size and creates a uniform grid. If X and Y are the lowest and highest values of the attribute, then the width W of intervals of recency or frequency will be,

$$W = (X - Y)/N \quad (2)$$

On the other hand, equal-depth divides the recency and frequency range into N intervals, each containing approximately the same number of samples. Apparently, equal-depth partitioning gives better data scaling and better handling of skewed data, especially if there is some outliers occurs in data distribution.

Both types of segmentation method are illustrated below.

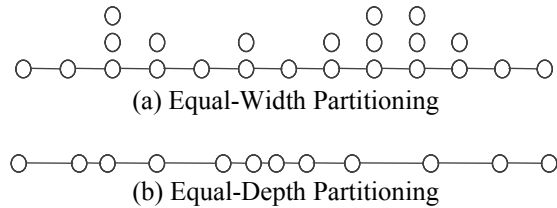


Figure 3. Types of Segmentation Method.

3.3 Transition Semantics

Moreover, in order to give semantic to the preference transition, there are three different type of transition based on RF table; “Stable”, “Rising”, and “Fading” transition. A “Stable” transition means that the correlation between two items are strongly connected because both items are in the same segment, while “Rising” denotes that the transition is moving up from a lower to an upper segment. The last type of transition is called “Fading” transition which means that the particular transition is weakening from an upper to a lower segment. The style of each type of transition’s semantic is shown in Figure 4.

By observing the type of transition as shown in Figure 4, the clarity of preference’s movement becomes easier to be spotted and observed. When the connection between two items is stable, it indicates that the two items are equivalent in term of both recency and frequency values. It is also implying that the users’ preferences are moving in the same segment of item. That is to say, there are no movements of users’ preferences to

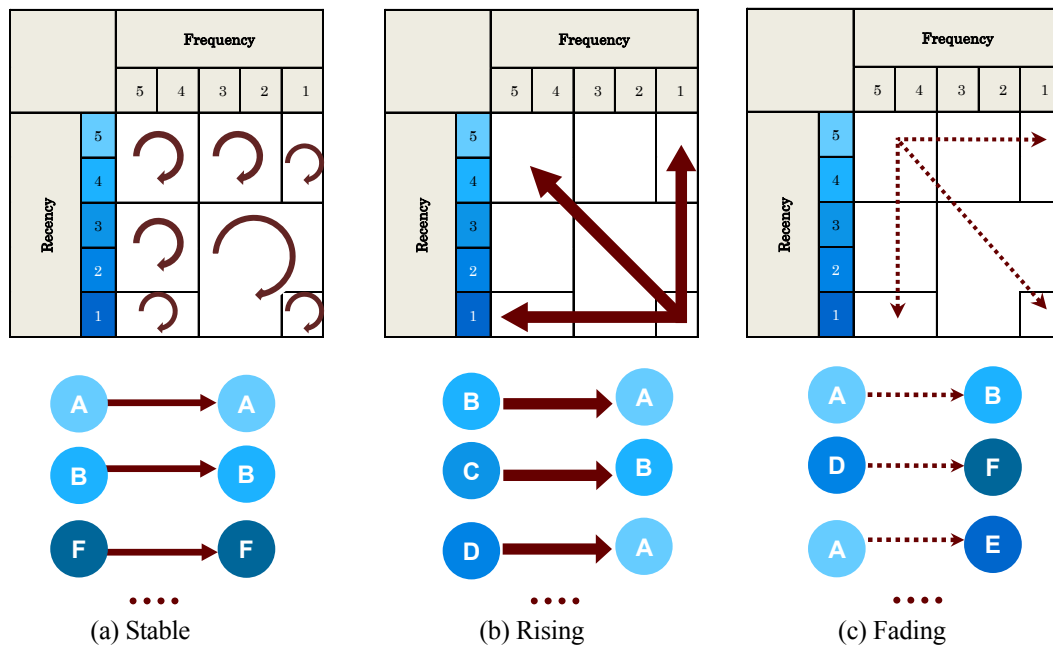


Figure 4. The Definition of Transaction Class.

another item from neither lower nor upper segment.

In addition, when the connection is rising from a lower segment item to an upper segment item then it clarifies that there is a transition of users' preferences to the item with higher recency and frequency value. On the other hand, a fading connection impresses the opposite implication, showing that such transition is outdated and infrequently occurs.

4. NUMERICAL EXPERIMENTATION

4.1 Preprocessing Stage

As sample data, a sales dataset of Northwind database (Microsoft), which consists of 1,902 transactions in two years span time, has been analyzed for experimentation. The detail of the Northwind dataset is shown in Table 3.

Table 3. Dataset details.

Number of users	97
Number of items	77
Number of transactions	1,902
Date range DD/MM/YY (transaction's timestamp)	04/07/96 06/05/98
Number of links	1,967
Preprocessing time	13m 21s

4.2 Result

As introduced in Section 3, the computation are being done in six stages, started by pairing every item that bought by every customers contained in transaction table. For every item pair, then the Simpson's coefficient and Mann-Whitney U-test are performed and the result of computations is stored in a separate table called "links" table. Also, in this step, the frequency of each transition is computed by counting the number of repeating item pairs that appear in links table.

The global date range of the transaction is between 04/07/96 and 06/05/98 in DD/MM/YY format. The recency score is taken from the latest purchase of each item, whereas the frequency score is determined from the transaction's volume of particular item. Afterward, the partitioning of recency and frequency classes, both for equal-width and equal-depth partition can be obtained. The following tables show the result of computations.

Table 4. Recency Partitions.

Recency Class	Equal-width partitioning		Equal-depth partitioning	
	Date interval (DD/MM/YY)		Date interval (DD/MM/YY)	
Class 1	04/07/96	15/11/96	04/07/96	13/03/98
Class 2	16/11/96	30/03/97	14/03/98	21/07/98
Class 3	31/03/97	12/08/97	22/07/98	21/04/98
Class 4	13/08/97	25/12/97	22/04/98	29/04/98
Class 5	26/12/97	06/05/98	30/04/98	06/05/98

Table 5. Frequency Partitions.

Frequency Class	Equal-width partitioning		Equal-depth partitioning	
Class 1	2	11	2	19
Class 2	11	19	19	24
Class 3	19	28	24	27
Class 4	28	37	27	33
Class 5	37	46	33	46

Table 6. RF-Table Segmentation Result.

		Frequency				
		5	4	3	2	1
Recency	5	18		5		0
	4	Hot items		Appearing items		New item
	3	10		40		0
	2					
	1	4				0
	Old items				Bad	

In fact, the chosen partition method is highly depends on the distribution of dataset because it will directly affect the RF-table segmentation. In Table 6, the generated RF-table based on equal-depth partitioning shows the segmentation of all items. The experiment result also has shown that 52% (40 of total items) of items are concentrated in “Fading” segmentation whereas there are no “New” or “Bad” items occur in the visualization.

For making details see clearly and RF-based method easily understand, a part of Figure 5 needs to be magnified. The part concerted item NO.15 which is specific node in this network has been choose to be magnified, because the transitions of users’ preferences for item 15 contain all three classes and also Item 15 delivered the user preference to another 10 items in four different item segments.

The general view of RF-based visualized preference transition is presented in Figure 5 with weight of Simpson’s coefficient is set to 0.5. Note that the smaller the Simpson’s coefficient is, the more complex its generated visualization. Magnified square expresses the state of items, for examples, item 13 is hot (category A) and item 3 is appearing (category B).

To be more clearly, the visualization can also be centered for the specific item. That is to say, the generated visualization is only focused to such particular item and its transitions to give more detaile’d information about the observed item. Figure 6 shows that the partial view of item 15 is divided into three types of visualization based on its transition. By doing so, the transitions of users’ preferences for item 15 can be clarified. Figure 6.a shows that there are five stable transitions from item 15 to the same segment of item (E-segment item). As shown in Figure 6(b), the item 15 also obtained 3 fading transition from two A-segment items and one D-seg-

ment item. Next, in Figure 6(c), it is shown that item 15 delivered the user preference to another 10 items in four different item segments, meaning that the users are changing their preference by choosing other newer and more popular items.

4.3 Discussion

Figure 7 demonstrates the comparison of the proposed method on generated visualization network based on recency and frequency with the original approach. The trends on choosing product items can be easily observed in RF-based visualization. This is mainly because in original scheme the frequency value is not considered to be occupied in preference analysis. As consequence, the original visualization suffers from trends detection

The experiment result has shown the potential of simplification introduced in Hayashi *et al.* (2010). Grouping the nodes where there is way to start at some node and follow a sequence of edges the eventually loops back to the node again, the preference transition becomes more clear. The simplified graph is called directed acyclic graph. For example, we have found four groups in the graph as shown in Table 7.

Table 7. Existing Groups (Cycles) in RF-Based (Simpson’s coefficient = 0.5).

Group	ItemID	Group	ItemID
G1	1	G2	8
	18		7
	22	G3	29
	37		66
	59	G4	32
	74		69

In addition, the generated visualization has involved all the three main properties of graph, which is node size, node color, and edge style. Here, the node size and color in RF-table have been used to determine whether an item is a strong (in upper segment of RF-table) or weak (in lower segment), and also the edge style has been able to describe the type of user movement on choosing product items. Its ability of detecting trends possibly could make another possibility to deliver an alternative visual recommendation system based on preference analysis.

5. CONCLUSION

This research has demonstrated a method for improving the visualized preference transition network used for visualizing customer movements/ trends in selecting items based on recency and frequency. The generated visualization has shown better result compared to

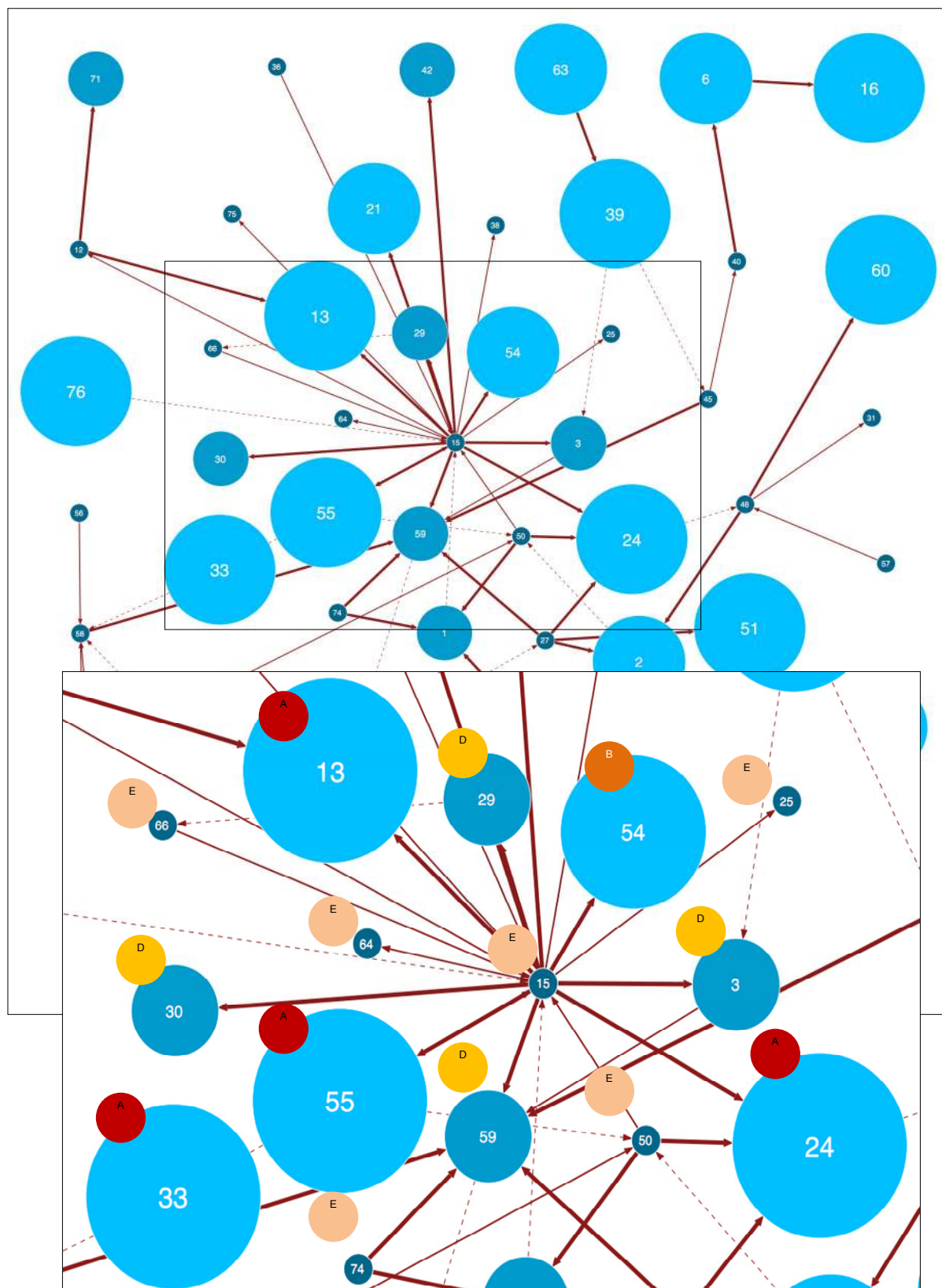
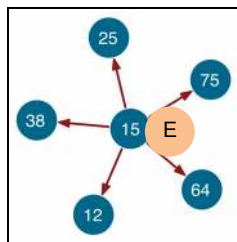
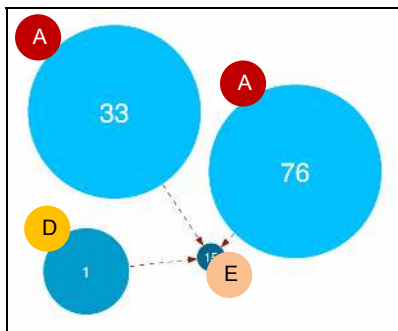


Figure 5. General View of RF-Based Visualized Preference Transition (Equal-depth partitioning, Simpson's coefficient = 0.5).

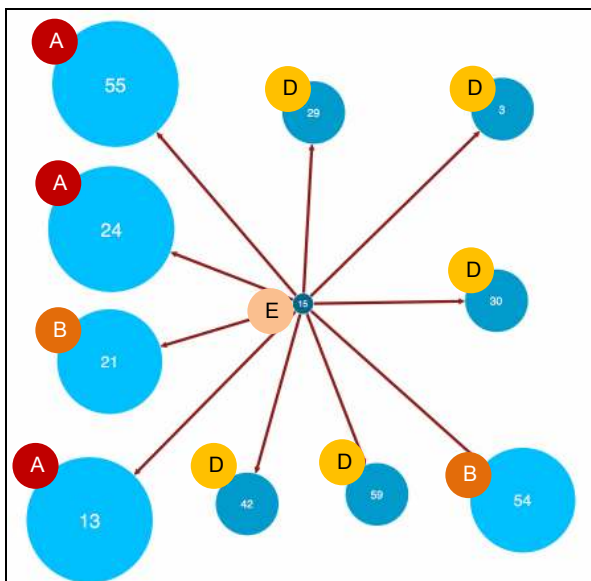
the original method in regards to easier observation of preference trends because the recency and frequency value were taken into account. Therefore, from the overall view of preference transition network, it can be clearly identified and recognized which items are strong or which are weak by observing the node size and color. Also, the movement of users' preferences on choosing product items is shown by the thickness of the edges.



(a) Stable Transition



(b) Fading Transition



(c) Rising transition

A = Hot item; B = Appearing item; D = Common item; E = Old item.

Figure 6. Partial Views of Item 15 (Equal-Depth Partitioning, Simpson's Coefficient = 0.5).

In future work, there are some opportunities to propose another approach in visual recommender system field based on preference analysis. It is also necessary to consider the evaluation as well as the optimization of preference transition algorithm. Moreover, the technique can be also applied to another domain, such as automobile brand selection transition and jewelry brand selection transition.

REFERENCES

Aczel, A. D. and Sounderpandian, J. (2005), Complete Business Statistics, McGraw-Hill.

Aggelis, V. and Christodoulakis, D. (2005), Customer Clustering using RFM Analysis, *9th WSEAS International Conference on Computers*.

D'Auria, T. (2009), The Fall of RFM Analysis, http://www.imn-unlocked.com/bi-creativecomputing/e_article001467796.cfm?x=b11,0,w.

Graphviz, <http://www.graphviz.org/>.

Han, J. and Kamber M. (2006), Data Mining: Concepts and Techniques, Elsevier Inc.

Hayashi, Y., Saga, R., and Tsuji, H. (2009), Competition State Visualization for Sales Record Mining, *Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 335-340.

Hayashi, Y., Masruri, F., Saga, R., and Tsuji, H. (2010), Enhanced Visualization on Preference Transition for Sales Records, *8th IEEE International Conference on Industrial Informatics (INDIN2010)*, 361-366.

Keim, D. A. (2002), Information Visualization and Visual Data Mining, *IEEE Trans. on Visualization and Computer Graphics*, **8**, 1-8.

Miglautsch, J. R. (2002), Thoughts on RFM scoring, *The Journal of Database Marketing*, 67-72.

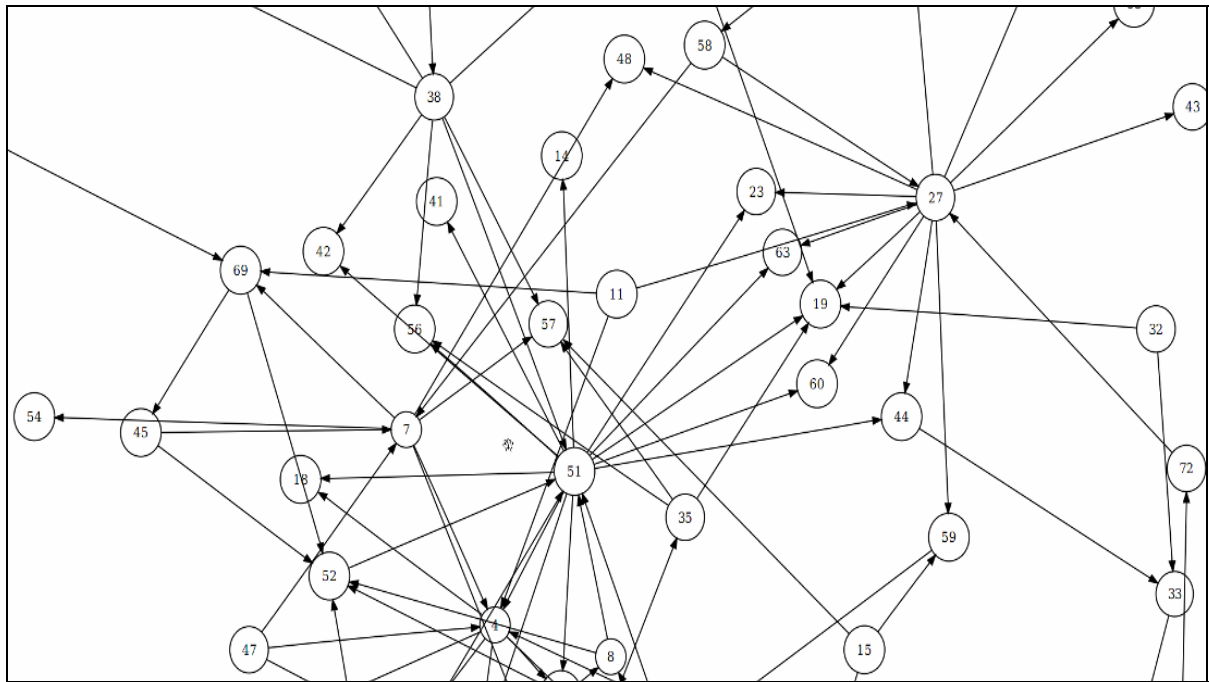
Northwind dataset, <http://www.microsoft.com/>.

Ohsawa, Y., Benson, N. E., and Yachida, M. (1998), Key Graph: Automatic Indexing by Cooccurrence Graph based on Building Construction Metaphor, *Proceedings. Advanced Digital Library Conference*, 12-18.

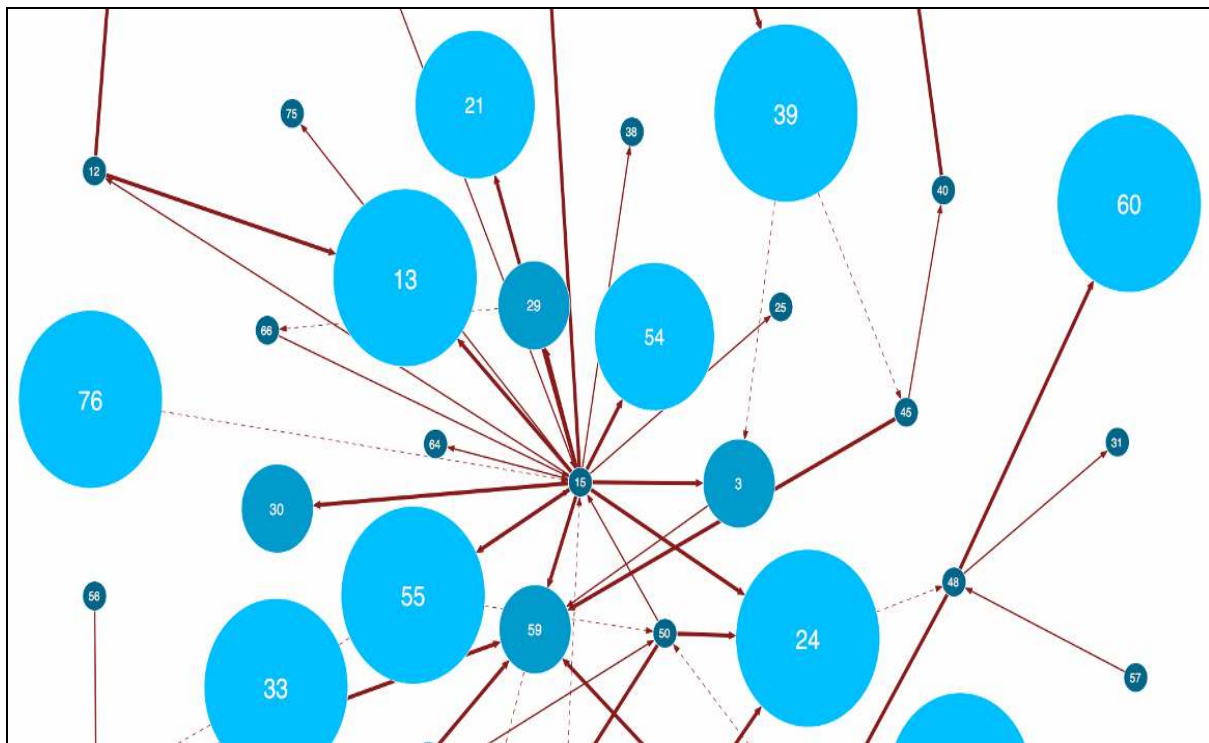
Saga, R., Terachi, M., Sheng, Z., and Tsuji, H. (2008), FACT-Graph: Trend Visualization by Frequency and Co-occurrence, *KI 2008. Lecture Notes on Artificial Intelligence, Springer-Verlag Berlin Heidelberg*, **5243**, 308-315.

Saga, R., Hayashi, Y., and Tsuji, H. (2008), Hotel Recommender System based on User's Preference Transition, *IEEE International Conference on Systems, Man and Cybernetics (IEEE/SMC 2008)*, 2437-2442.

Terachi, M., Saga, R., Sheng, Z., and Tsuji, H. (2008), Visualized Technique for Trend Analysis of News Articles, *Lecture Notes on Artificial Intelligence, Springer-Verlag Berlin Heidelberg*, **5027**, 659-668.



(a) Original Visualization



(b) RF-based Visualization (Equal-Depth Partitioning)

Figure 7. Comparison of Visualized Preference Transition (Simpson's coefficient = 0.5).